

A determination of the spatial concordance between Lyme disease incidence and habitat probability of its primary vector *Ixodes scapularis* (black-legged tick)

Samuel F. Atkinson¹, Sahotra Sarkar², Aldo Aviña³, Jim A. Schuermann⁴, Phillip Williamson⁵

¹Institute of Applied Science, Department of Biological Sciences, University of North Texas, Denton, USA;

²Integrative Biology and Philosophy, University of Texas, Austin, USA; ³Department of Environmental and Occupational Health, School of Public Health, Health Science Center, University of North Texas, Fort Worth, USA; ⁴Infectious Disease Control Unit, Texas Department of State Health Services, Austin, USA; ⁵Scientific Affairs, Creative Testing Solutions, Tempe, USA

Abstract. The spatial distribution of *Ixodes scapularis*, the most common tick vector of the bacterium *Borrelia burgdorferi*, the cause of Lyme disease in humans, has not been studied previously in Texas, United States of America. It has only rarely been reported in this state, so its local, spatial relationship to the distribution of this disease is unknown. From an epidemiological perspective, one would tend to hypothesise that there should be a high degree of spatial concordance between habitat suitability for the tick and incidence of the disease. Both maximum-entropy modelling of the tick's habitat probability and modelling of human incidence of Lyme disease using spatially adaptive filters provide reliable portrayals of the spatial distributions of these phenomena. Even though rates of human cases of Lyme disease as well as rates of *Ixodes* ticks infected with *Borrelia* bacteria are both relatively low in Texas, the best data currently available indicate that the assumption of high levels of spatial concordance would not be correct in Texas (Kappa coefficient of agreement = 0.039). It will take substantially more data to provide conclusive findings and to understand the results reported here, but this study provides an approach to begin understanding the discrepancy.

Keywords: Lyme disease, *Ixodes scapularis*, habitat probability, United States of America.

Introduction

According to the Centers for Disease Control and Prevention (CDC) in the United States of America (USA), Lyme disease is the most common, tick-borne disease with over 20,000 confirmed new cases reported each year (CDC, 2007). This number continues to increase, underscoring the need for targeted prevention strategies, early disease recognition and treatment and a sustainable surveillance system (Bacon et al., 2008). The median incidence rate of Lyme disease among the 50 states in USA in 2009 was 0.5 cases per 100,000 population each year. However, the national mean incidence rate was reported as 13.4 cases per 100,000, indicating that the data are highly skewed by a few states with exceptionally high rates (only 13 states had incidence rates above the national mean rate, seven of

which had rates exceeding 50 cases per 100,000 population). Texas experiences Lyme disease incidence at almost exactly the national median rate (0.4 cases per 100,000) with 23 states reporting higher rates and 22 states reporting lower rates (DVBID, 2010). Since the year 2000, the incidence of Lyme disease in Texas has normally varied between 50 and 100 confirmed cases per year (Fig. 1). Based on improved serologic testing and a modified laboratory confirmation (pathogen isolation from a clinical specimen or demonstration of diagnostic levels of IgM or IgG antibodies in serum or cerebrospinal fluid), the case definition of Lyme disease changed in the beginning of 2009. That year, Texas reported 88 confirmed cases and 188 additional probable cases. In 2010, there were only 52 confirmed cases and 90 additional probable cases.

The black-legged tick, *Ixodes scapularis*, is reported as the primary (but not only) vector of *Borrelia burgdorferi*, the bacterium that causes Lyme disease in humans (Dennis et al., 1998). We have recently reported (Williamson et al., 2010) the results of a 4-year study across Texas, where a total of 903 ticks, representing 11 tick species, which had been submitted to the Tick-Borne Disease Research Laboratory at the University of North Texas Health Science Center

Corresponding author:

Samuel F. Atkinson
Institute of Applied Science
Department of Biological Sciences, University of North Texas
1155 Union Circle #310559, Denton, TX 76203, USA
Tel. +1 940 565-2694; Fax +1 940 565-4297
E-mail: atkinson@unt.edu

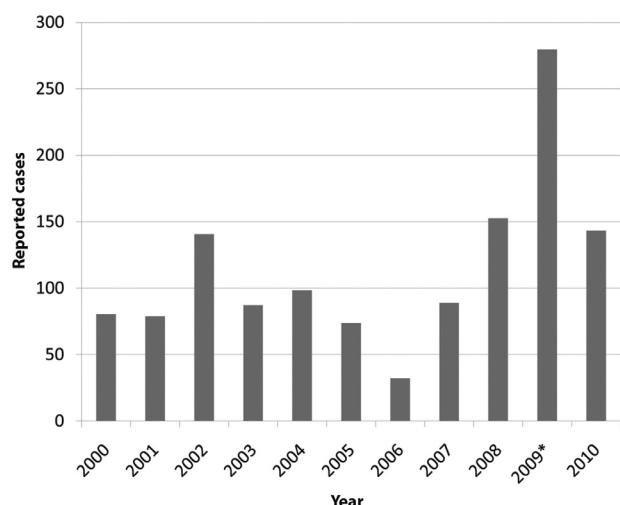


Fig. 1. Reportable cases of Lyme disease in Texas, 2000-2010.

*Case definition of confirmed cases changed in the beginning of 2009.

(UNTHSC) for analysis of both tick identification and rates of *Borrelia* infection. These ticks had been attached to and removed from, humans and were subjected to molecular analysis for detection of *Borrelia* spp. bacteria. A total of 0.8% of *Amblyomma*, *Dermacentor* and *Ixodes* ticks in that study were infected with *Borrelia* bacteria.

Because neither tick presence nor Lyme disease incidence is evenly spread across Texas, we became interested in understanding the spatial concordance between the two types of data. Based on a technique that we developed for understanding the spatial concordance between Rocky Mountain spotted fever and the American dog tick *Dermacentor variabilis* (Atkinson et al., 2012), the study reported here compared two maps of Texas related to Lyme disease. The first is based on environmental conditions and represents the probability of suitable habitat for *Ixodes* ticks (habitat probability mapping), while the second represents the incidence rate of Lyme disease across the state (disease incidence mapping). From an epidemiological perspective, one would tend to hypothesise that there should be a high degree of spatial concordance between these two maps. Even though rates of human cases of Lyme disease in Texas and rates of infected ticks in Texas are both relatively low, the best data currently available indicate that the assumption of high levels of spatial concordance cannot be correct. We have investigated this, but since the results presented here must be considered as preliminary, it will take substantially more data to provide conclusive findings.

Material and methods

Our research involved mapping the probability of occurrence that the disease vector exists in the environment (i.e. tick habitat probability mapping), mapping the incidence of Lyme disease in the human population (disease incidence mapping) and examining the spatial concordance between the probability map and incidence map.

Habitat probability mapping

Because the *Ixodes* ticks are the primary vector of the bacterium *Borrelia*, its presence will increase the risk of Lyme disease (Glass et al., 1995; Kitron and Kazmierczak, 1997; Eisen et al., 2006). However, public health efforts to determine the spatial distribution of tick populations in Texas (and in many other parts of the world) are limited due to insufficient data collection (Dennis et al., 1998). Fortunately, ticks have certain environmental requirements for survival (Glass et al., 1995; Brownstein et al., 2003), so the potential distribution of tick populations can be estimated with species distribution models that allow mapping habitat probability (that is the probability of occurrence) based on presence-only occurrence data and the spatial distribution of environmental variables.

Tick habitat maps can be thought of as models of the probability of an area having suitable habitat for the ticks (dependent variable) based on a series of environmental covariates (independent variables). Following a well-established risk assessment protocol in the epidemiological literature (Peterson, 2008; Gonzalez et al., 2010; Sarkar et al., 2010), these maps can then be used as a surrogate, or indicator, of the risk of encountering *Ixodes* ticks. Previous literature (Glass et al., 1995; Brownstein et al., 2003) suggests that the environmental covariates most indicative of *Ixodes* ticks are: temperature (maximum, minimum and mean), moisture, soil types and forest cover.

The study reported here utilised most of the environmental covariants mentioned (soil type was investigated but was removed from the model since it had no influence on the resultant map). The environmental covariants, in conjunction with tick occurrence data, were modelled using a maximum-entropy (MaxEnt) algorithm incorporated in the software package of that name. MaxEnt works well with limited, presence-only datasets and quickly processes data to create probability distributions based on environmental parameters and species occurrence (Phillips et al., 2006; Phillips and Dudík, 2008). Among the several

alternative protocols available for species distribution mapping (Franklin and Miller, 2009), MaxEnt was chosen for this study because a comprehensive comparative assessment found its predictive performance to be consistently among the highest performing methods (Elith et al., 2006). MaxEnt estimates a distribution across geographic space (Phillips et al., 2006; Phillips and Dudík, 2008) or, equivalently, probability densities in covariate space (Elith et al., 2006).

MaxEnt requires the geographic coordinates of the locations where the species of interest has been observed as input (i.e. the presence sites). Next, it examines the data for each environmental covariate at each of those locations (this is termed “conditional density of the covariates at the presence sites”). Then, for a series of randomly selected locations in the study area called background sites (which may or may not contain some of the presence sites), the data for each environmental covariate (called “marginal density of covariates across the study area) are examined. The model computes a metric called the area under the curve (AUC) value, which measures the quality of a ranking of sites’ probability to support the species under consideration (Fielding and Bell, 1997). The AUC is the probability that a randomly chosen presence site will be ranked above a randomly chosen absence site. A random ranking has, on average, an AUC of 0.5 and a perfect ranking achieves the best possible AUC of 1.0. Models with values above 0.75 are considered useful for understanding the distribution of the species in question (Elith, 2000).

The tick occurrence data were generated between 2004 and 2010 from submissions to the Texas Department of State Health Services (TDSHS). Testing of the TDSHS samples by the UNTHSC, an ongoing endeavour, so the study reported here not only includes the 903 ticks collected between 2004 and 2008 in our original study (Williamson et al., 2010), but includes also 555 additional tick submissions col-

lected between 2008 and 2010. The occurrence data were aggregated by TDSHS to the zip-code level.

Molecular screening for tick-borne pathogens, including *B. burgdorferi*, was performed at UNTHSC using the method of Williamson et al. (2010). These data were spatially tagged with the zip code of where the tick was encountered. For modelling purposes, it is unrealistic to assume that for zip codes with multiple tick occurrences, each human-tick encounter occurred at the exact same location (e.g. the spatial centroid of the zip code area). Because MaxEnt requires point data (a specific longitude/latitude or X/Y point coordinate based on various projection systems), the tick occurrence data were re-assigned X/Y locations by randomly distributing individual tick occurrences within the corresponding zip code. This allowed tick occurrences to be more realistically distributed across the entire zip code, rather than just at its centroid. The occurrence data were randomly distributed across the zip codes 30 times with each resultant distribution processed by MaxEnt. Those results were averaged to produce our habitat probability map and ultimately sampled on a 500 x 500 m grid to match the disease incidence mapping resolution and thus allow an investigation in the spatial concordance between the two datasets.

The environmental parameters entered into our MaxEnt model are shown in Table 1 (National Elevation Dataset, 2010; National Land Cover Dataset, 2010; PRISM Climate Group, 2010; U.S. Department of Agriculture, 2010; WorldClim.org, 2010). These parameters were chosen based on the studies discussed earlier, which identified several key ecological variables associated with the presence of *Ixodes* ticks. By mapping the distribution of these ticks at a geographical point scale, something which has only been done at such a large scale in one previous study (Atkinson et al., 2012), we have developed a surrogate of Lyme disease exposure risk based on locations and attendant environmental data associated with tick submissions.

Table 1. Data, sources and reasoning for inclusion in modelling.

Data	Source	Reasoning
Elevation	National Elevation Dataset	Determination if elevation is a factor of tick presence in Texas
Temperature	WorldClim.org	Ticks sensitive to temperature extremes
Relative humidity	PRISM Climate Group, calculated from dew point and ambient temperature*	Tick presence/activity is dependent on water vapour
Land use/land cover	National Land Cover Dataset	Forest cover relates to tick probability
Soil texture	Soil Survey Geographical Database	Soil texture influences moisture content

*Dew point is the temperature at which the air can no longer adsorb water vapor such that water vapor will condense into liquid water. PRISM provides data on both dew point and temperature, and therefore relative humidity can be solved for analytically, which has been reported as the best “moisture” environmental parameter for predicting tick habitat.

Disease incidence mapping

The second aspect of our study was to develop a map expressing Lyme disease incidence in Texas, which is reported at the zip code level. When the number of cases are normalised by population size in each zip code to report incidence rates of Lyme disease, the data may be susceptible to instability due to small population sizes in some zip code areas, i.e. the “small numbers problem” (MacEachren et al., 1998; Jerrett et al., 2010). In order to reduce the influence of this problem, we smoothed the disease count data using kernel density estimation methods. These methods compute rates for overlapping circular areas, using spatial filters centered on a regular grid of points, which allows the normalisation of incidence rates that would otherwise be over- or under-estimated, creating a more accurate representation of disease incidence (Rushton et al., 2004; Tiwari and Rushton, 2005). More detailed discussions of spatial filters are available (Brillinger, 1994; Kafadar, 1994) and several illustrations of the method can be found (Lai et al., 2004; Yang et al., 2006; Kloog et al., 2009). Here, our spatial filters were centered on a 10-mile grid superimposed on the state, increasing in size by capturing surrounding zip code areas when the centroid of a zip code fell within the kernel until a minimum population size of 10,000 of all captured zip codes was reached. This approach results in kernels of varying size based on the underlying population density and is referred to as spatially adaptive filtering. Larger-sized filters occur in sparsely populated areas and smaller ones occur in densely populated areas. Larger filters provide more smoothing but also result in loss of geographical detail whereas smaller filters provide

greater geographical detail. Because adaptive filtering is based on varying filter size, resultant maps provide high geographical resolution in those areas where such detail is expected, while maintaining rate stability in rural areas with sparse populations (Tiwari, 2011).

The number of Lyme disease cases in all zip code areas captured by the spatial filter and the total population in those zip codes was used to generate a more representative incidence rate across Texas. Because each grid point in the study area now had a calculated incidence rate, incidence rates could be interpolated across Texas using the inverse distance weighted (IDW) method (Tiwari and Rushton, 2005). Ultimately, the map was resampled into 500 x 500 m grid cells for spatial concordance analysis.

Spatial concordance mapping

The spatial agreement between our habitat probability and disease incidence maps was measured using an overlay technique where categorical data from two input maps generate a third categorical data map for which each output category represents every combination of the two input categories. The technique, originally developed to detect and quantify land use change (Serra et al., 2003; Jensen, 2005), compares the assigned categories from each input map at a given location (i.e. a pixel) and uses a decision matrix such as the one shown in Table 2 to assign the output class for that same location in the output map. Because the technique provides a measure of concordance and also produces a derived map, it offers both visual and quantitative expressions of map agreement.

Here, the categories in the habitat probability map,

Table 2. Spatial concordance matrix, consisting of habitat and incidence classes for N pixels.

Spatial concordance	Habitat probability map class					Row total
	Lowest 20% (1)	2 nd quintile (2)	3 rd quintile (3)	4 th quintile (4)	Highest 20% (5)	
Incidence rate map class	Lowest 20% (1)	1 _{h1, i1}	2 _{h2, i1}	3 _{h3, i1}	4 _{h4, i1}	5 _{h5, i1} $\Sigma_{row\ 1}$
	2 nd quintile (2)	6 _{h1, i2}	7 _{h2, i2}	8 _{h3, i2}	9 _{h4, i2}	10 _{h5, i2} $\Sigma_{row\ 2}$
	3 rd quintile (3)	11 _{h1, i3}	12 _{h2, i3}	13 _{h3, i3}	14 _{h4, i3}	15 _{h5, i3} $\Sigma_{row\ 3}$
	4 th quintile (4)	16 _{h1, i4}	17 _{h2, i4}	18 _{h3, i4}	19 _{h4, i4}	20 _{h5, i4} $\Sigma_{row\ 4}$
	Highest 20% (5)	21 _{h1, i5}	22 _{h2, i5}	23 _{h3, i5}	24 _{h4, i5}	25 _{h5, i5} $\Sigma_{row\ 5}$
Column total		$\Sigma_{col\ 1}$	$\Sigma_{col\ 2}$	$\Sigma_{col\ 3}$	$\Sigma_{col\ 4}$	$\Sigma_{col\ 5}$ N

represented by columns in the decision matrix that show variation from pixels with the lowest 20% of tick habitat probability (column 1) through to the highest (column 5) with classes 2, 3 and 4 representing the three middle probability classes. A similar categorisation for the Lyme incidence rate was utilised, where the rows vary in the same way from pixels with the lowest 20% incidence rates through to the highest rates. These two input categorised maps were overlaid producing a resultant output map containing 25 possible classes: these two input categorised maps were overlaid producing a resultant output map containing 25 possible classes: class 1 represents pixels with the lowest 20% incidence rate and the lowest 20% habitat probability; class 2 represents pixels with the lowest 20% incidence rate and the second lowest habitat probability; and so on through class 25 representing pixels with the highest incidence and highest habitat probability.

Overall concordance is simply the sum of the count of pixels along the major diagonal of the decision matrix, divided by the total number of pixels analysed. If two maps were in perfect spatial concordance, all pixels would fall on the major diagonal of the decision matrix (in this case 1, 7, 13, 19 and 25), resulting in a concordance of 1.0. Perfect spatial concordance is rare and every non-concordant pixel falls in one of the off-diagonal cells in the decision matrix. The column and row totals around the margin of the decision matrix (referred to as marginals) can be used to compute the concordance for each class. Finally, the Kappa coefficient of agreement (K) provides a measure of the amount of agreement between two maps and accounts for chance or random agreement between the two maps (Congalton et al., 1983):

$$K = \frac{N \sum x_d - \sum (x_r \times x_c)}{N^2 - \sum (x_r \times x_c)} \quad (\text{equation 1})$$

where N = total number of pixels analysed; x_d = count on the diagonal for row i and column b ; x_r = total count on row i ; x_c = total count on column b .

The conditional error in agreement between the likelihood of a Lyme disease incidence class correctly falling into the appropriate habitat probability class (e.g. likelihood that Lyme disease incidence class 1 will fall into *Ixodes* habitat probability class 1) can be calculated as:

$$E_L = 1 - \frac{x_d}{x_c} \quad (\text{equation 2})$$

where E_L = conditional error for *Ixodes* habitat probability class i falling into Lyme incidence class i .

Likewise, the conditional error in agreement between the likelihood of an *Ixodes* habitat probability class correctly falling into the appropriate Lyme disease incidence class (e.g. likelihood that *Ixodes* habitat probability class 1 will fall into Lyme disease incidence class 1) can be calculated as:

$$E_H = 1 - \frac{x_d}{x_r} \quad (\text{equation 3})$$

where E_H = conditional error for Lyme incidence class i falling into *Ixodes* habitat probability class i .

Results

Texas has a land area of approximately 678,000 km² and a population exceeding 25 million. This overall density (approximately 37 people per km²) is not evenly divided across the state - nearly a third of the state's population lives in 10 cities. Additionally, the state is divided into a fairly distinct eastern and western climatic regime, where the eastern half tends to have more potential precipitation than potential evapotranspiration, while the western half tends to be the opposite. These two characteristics appear to have a substantial influence on the maps generated for this research.

Habitat probability map

Fig. 2 shows the MaxEnt output of the estimated distribution of *Ixodes* ticks, based on environmental characteristics modelled from the collection of ticks between 2004 and 2010. The areas of highest probability occur in eastern and north-eastern Texas. The AUC number of the map in Fig. 2 is 0.843 and therefore the habitat probability map for *Ixodes* ticks can be considered predictively adequate based upon Elith et al. (2006) and Pawar et al. (2007) criteria. The map mimics intuition in the sense that the dry conditions and sparse vegetation in the western half of the state should limit tick survivability, which is corroborated by the results of the MaxEnt modelling. Table 3 indicates that nearly 95% of Texas' land area is predicted to have less than a 50% probability of having suitable habitat for *Ixodes* ticks. Less than 0.5% of Texas has a 75% or greater probability of having suitable habitat, yet Texas' incidence rate is almost exactly the nation's median rate for all states.

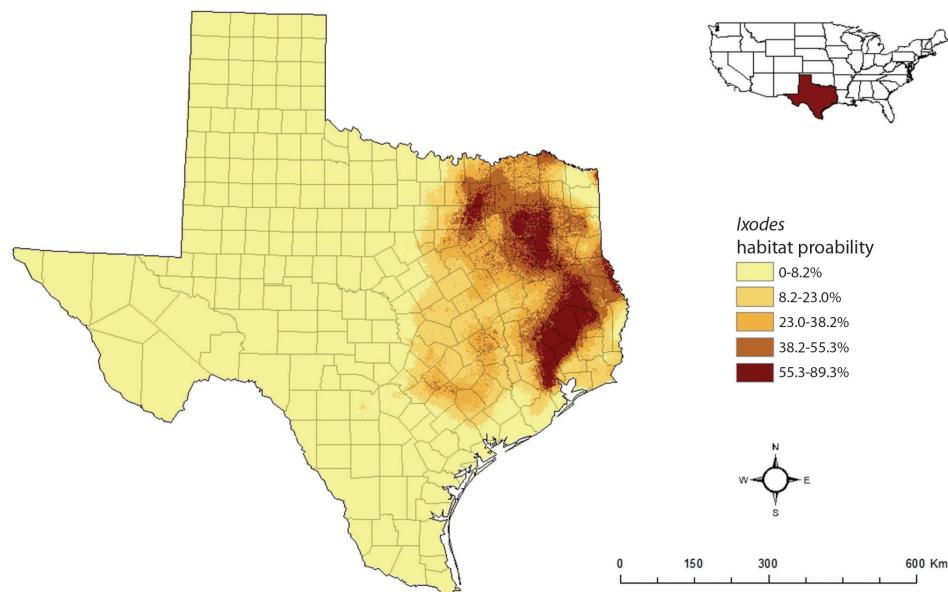


Fig. 2. Estimated distribution of the *Ixodes scapularis* (black-legged tick) based on ticks collected between 2004 and 2010 and MaxEnt modelling of habitat probability.

Table 3. MaxEnt modelling results for *Ixodes* tick habitat in Texas.

Suitable-habitat probability	Landmass percentage	Cumulative percent
0% to 5%	70.9%	70.9%
>5% to 10%	3.6%	74.6%
>10% to 15%	3.0%	77.6%
>15% to 20%	3.6%	81.2%
>20% to 25%	2.9%	84.1%
>25% to 30%	2.5%	86.6%
>30% to 35%	2.5%	89.1%
>35% to 40%	2.2%	91.3%
>40% to 45%	1.9%	93.2%
>45% to 50%	1.7%	94.9%
>50% to 55%	1.4%	96.2%
>55% to 60%	1.3%	97.5%
>60% to 65%	0.9%	98.4%
>65% to 70%	0.8%	99.2%
>70% to 75%	0.5%	99.7%
>75% to 80%	0.2%	99.9%
>80% to 85%	0.1%	100.0%
>85% to 90%	0.0%	100.0%
>90% to 95%	0.0%	100.0%
>95% to 100%	0.0%	100.0%

Lyme disease incidence map

Over the past decade the incidence of Lyme's disease across Texas ranges from zero to 12.6 cases per 10,000 people. Fig. 3 presents the map of Lyme disease incidence rates in Texas. The highest burdens occur in central Texas, away from the most densely populated cities but also away from areas of high tick habitat probability. There are a few other areas in Texas with relatively high incidence rates, for example on the northern border with Oklahoma and the southern border with Mexico, but all of these areas are removed from the most densely populated and large cities in Texas. Nearly 85% of the state's area experienced a rate of less than 1 case of Lyme disease per 10,000 since 2000 (Table 4). Only 1.6% of the state's area experienced a rate higher than 5 cases per 10,000.

Concordance map

Fig. 4 shows the result of overlay mapping between *Ixodes* tick habitat probability mapping and Lyme incidence rates, while Table 5 summarises the overall concordance between these two maps. Fig. 4 only shows (in black) where there was concordance between the top 20% of the state with the highest habitat probability and the top 20% of the state with the highest Lyme disease incidence. While 20% of the state's area is approximately 135,600 km², only 21,000 km², or 3.1% of the state has both high incidence of Lyme dis-

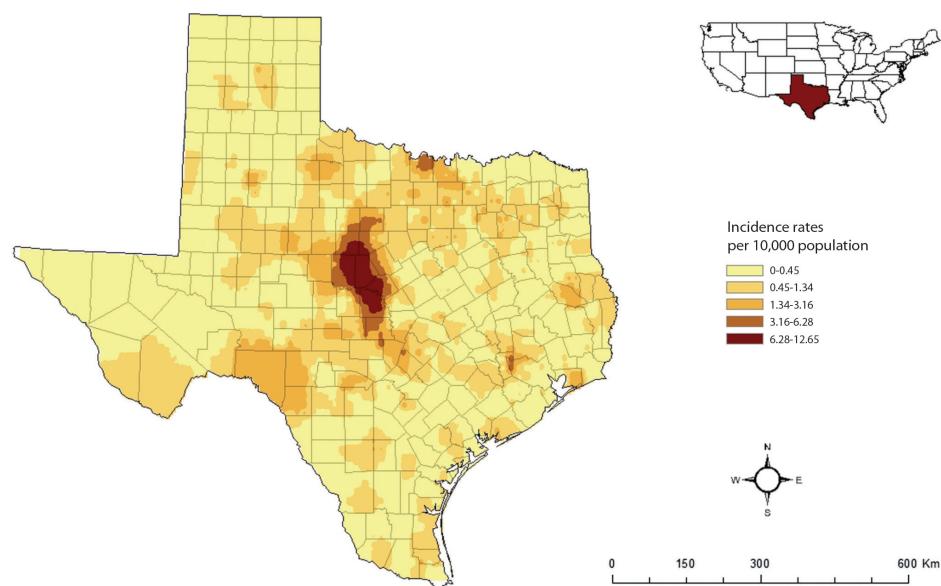


Fig. 3. Rates of Lyme disease based on cumulative cases from 2000-2010.

Table 4. Lyme disease incidence in Texas, 2000 to 2010 based on the spatially adaptive filter model.

Incidence rate per 10,000 population	Landmass percentage	Cumulative percent
0 to 0.5	61.5%	61.5%
>0.5 to 1.0	23.2%	84.7%
>1 to 1.5	6.1%	90.8%
>1.5 to 2.0	3.8%	94.6%
>2.0 to 2.5	1.8%	96.4%
>2.5 to 3.0	0.7%	97.1%
>3.0 to 3.5	0.4%	97.6%
>3.5 to 4.0	0.4%	97.9%
>4.0 to 4.5	0.3%	98.2%
>4.5 to 5.0	0.2%	98.4%
>5.0 to 5.5	0.2%	98.7%
>5.5 to 6.0	0.2%	98.8%
>6.0 to 6.5	0.1%	98.9%
>6.5 to 7.0	0.1%	99.0%
>7.0 to 7.5	0.2%	99.2%
>7.5 to 8.0	0.2%	99.4%
>8.0 to 8.5	0.3%	99.7%
>8.5 to 9.0	0.1%	99.8%
>9.0 to 9.5	0.0%	99.8%
>9.5 to 10.0	0.0%	99.9%
>10.0 to 10.5	0.0%	99.9%
>10.5 to 11.0	0.1%	100.0%
>11.0 to 11.5	0.0%	100.0%
>11.5 to 12.0	0.0%	100.0%
>12.0	0.01%	100.0%

ease and high habitat probability for *Ixodes* ticks. Table 3 summarises overall and class specific concordance for the state.

The overall concordance of these two maps is only 23.1% and nearly 40% of the overall concordance falls in those areas with both the lowest habitat probability and the lowest incidence rates (9.1%). The Kappa statistic, which removes chance agreement, indicates that there is less than 4% agreement between the two maps ($K = 0.039$). Kappa values between 0.40 and 0.80 are considered to represent moderate agreement, but less than 0.4 is considered to represent poor agreement (Landis and Koch, 1977).

The conditional error in agreement for a Lyme disease incidence class falling into the appropriate *Ixodes* tick habitat probability class ranges from a low of 57.5% error (Lyme incidence and habitat probability classes 1) to a high of 87.9% error (Lyme incidence and habitat probability classes 2). The conditional error in agreement for an *Ixodes* tick habitat probability class falling into the appropriate Lyme disease incidence class ranges from 57.3% error (habitat probability and Lyme incidence classes 1) to a high of 87.4% error (habitat probability and Lyme incidence classes 2).

Discussion

The reasons for low spatial concordance between areas exhibiting high Lyme disease incidence rates and high *Ixodes* tick habitat probability in Texas are cur-

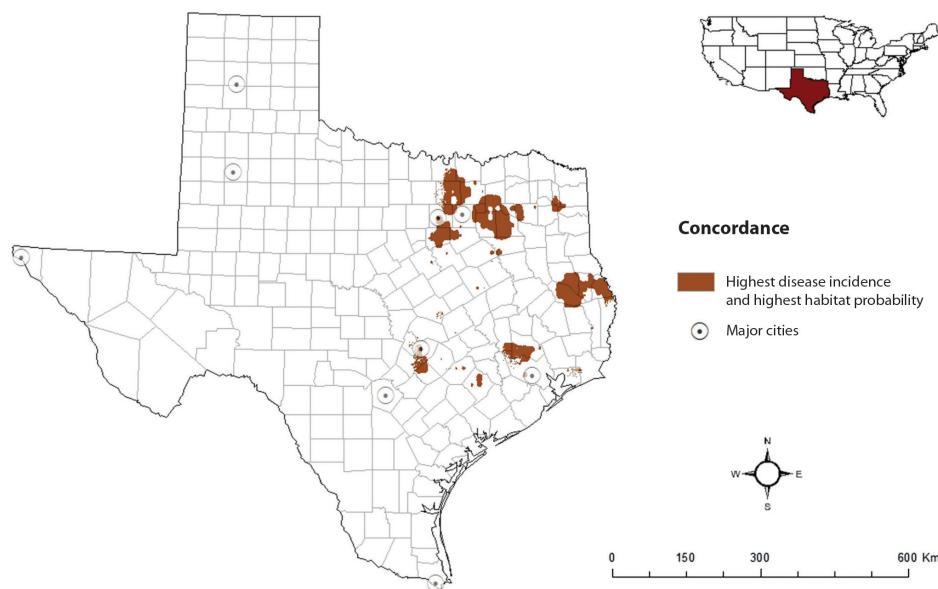


Fig. 4. Areas in red (approximately 3.1% of Texas' landmass) represent the overlap of Texas' highest 20% of *Ixodes* habitat probability and Texas' highest 20% of Lyme disease incidence rates ($K = 0.039$).

rently unknown. It may be partially due to various human populations' accessibility to hike and bike trails, greenbelts and campgrounds - areas where people may be exposed to more ticks than in less "wilderness-like" areas. Additionally, individuals testing positive for Lyme disease may have acquired the disease outside their county of residence. Another reason may be that areas that would normally not be suitable for *Ixodes* ticks according to MaxEnt modelling may actually be suitable due to agricultural and irrigation processes that change the landscape in a way that creates suitable moisture and vegetation regimes for the

ticks. These agricultural processes were not incorporated in the tick habitat model. Also, Martinez et al. (1999) discovered *B. burgdorferi* in deer in north-eastern Mexico. Lyme disease cases may be transported from (or to) Mexico, affecting Lyme disease incidence estimates in the USA-Mexico border region. However, more localised research must be done to accurately assess the origins of Lyme disease risk in specific regions.

This should be considered an initial investigation based on a relatively small dataset. With that caveat, the preliminary data seem to reveal that areas consid-

Table 5. Spatial concordance matrix of Lyme disease incidence by *Ixodes* tick habitat probability across Texas.

Spatial concordance	Habitat probability map class					Row total
	Lowest 20% (1)	2 nd quintile (2)	3 rd quintile (3)	4 th quintile (4)	Highest 20% (5)	
Incidence rate map class	Lowest 20% (1)	9.1%	4.8%	3.7%	1.5%	2.3% 21.4%
	2 nd quintile (2)	2.5%	2.3%	3.9%	4.5%	5.2% 18.5%
	3 rd quintile (3)	2.3%	2.9%	4.4%	5.2%	5.2% 20.1%
	4 th quintile (4)	3.3%	4.0%	4.1%	4.1%	4.5% 20.0%
	Highest 20% (5)	3.0%	5.1%	4.0%	4.8%	3.1% 20.0%
Column total		20.3%	19.1%	20.1%	20.2%	20.3% 100.0%

Percentages in the table represent what proportion of the total landmass of Texas (678,000 km²) falls within each pair of map classes.

ered to have an increased risk of encountering *Ixodes* ticks due to more suitable habitat are located in the eastern portions of Texas, while the pattern of actual Lyme disease incidence is more prevalent in central Texas. As of now, we speculate that this disparity may be due to insufficient tick sampling, that knowledge of tick submission programme in Texas may be unequally distributed across the state; that people travelling from low tick habitat probability areas to areas with high tick habitat probability, where they become exposed to the disease and/or a small possibility; that *B. burgdorferi* is being spread by another species of tick in this part of USA. Further, some anecdotal information suggests that south Texas has a relatively high deer population and incorporating deer density data into MaxEnt modelling may provide additional insights. The areas that we have shown to have a high degree of concordance for high disease incidence rates and high tick habitat probability (see Fig. 4) should be targeted for more focused research.

Adding more accurate tick collection data (e.g. field sampling programmes designed to precisely record the location of tick collection), more precise case location data, measuring human accessibility to areas at risk for tick infestation and perhaps deer density data will help in understanding why there exists a spatial disagreement between *Ixodes* distribution and Lyme disease incidence in Texas. Nevertheless, these results indicate that even with minimal data, by utilising advanced mapping techniques many more observations can be made about Lyme disease in Texas than would have been previously possible. By implementing a spatially adaptive filters analysis to human disease incidence and utilising MaxEnt habitat probability modelling, it has been possible to provide some initial, unexpected observations about Lyme disease in Texas.

References

- Atkinson SF, Sarkar S, Aviña A, Schuermann JA, Williamson P, 2012. Modelling spatial concordance between Rocky Mountain spotted fever disease incidence and habitat probability of its vector *Dermacentor variabilis* (American dog tick). *Geospat Health* 7, 91-100.
- Bacon RM, Kugeler KJ, Mead PS, 2008. Surveillance for Lyme disease - United States, 1992-2006. Available at: <http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5710a1.htm> (accessed on April 2011).
- Brillinger DR, 1994. Examples of scientific problems and data analyses in demography, neurophysiology and seismology. *J Comput Graph Stat* 3, 1-22.
- Brownstein JS, Holford TR, Fish D, 2003. A climate-based model predicts the spatial distribution of the Lyme disease vector *Ixodes scapularis* in the United States. *Environ Health Perspect* 111, 1152-1157.
- CDC, 2007. Lyme disease - United States, 2003-2005. Available at: <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5623a1.htm> (accessed on April 2011).
- Congalton RG, Oderwald RG, Mead RA, 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogramm Eng Remote Sensing* 49, 1671-1678.
- Dennis DT, Nekomoto TS, Victor JC, Paul WS, Piesman J, 1998. FORUM reported distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the United States. *J Med Entomol* 35, 629-638.
- DVBID, 2010. Lyme disease statistics. Online, statistics reviewed on September 13, 2010. Available at: http://www.cdc.gov/ncidod/dvbid/lyme/ld_statistics.htm (accessed on April 2011).
- Eisen RJ, Lane RS, Fritz CL, Eisen L, 2006. Spatial patterns of Lyme disease risk in California based on disease incidence data and modeling of vector-tick exposure. *Am J Trop Med Hyg* 75, 669-676.
- Elith J, 2000. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: Quantitative methods for conservation biology. Ferson S, Burgman M (eds). Berlin: Springer, 39-58 pp.
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129-151.
- Fielding AH, Bell JF, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24, 38-49.
- Franklin J, Miller JA, 2009. Mapping species distributions: spatial inference and prediction. Cambridge: Cambridge University Press.
- Glass GE, Schwartz BS, Morgan JM III, Johnson DT, Noy PM, Israel E, 1995. Environmental risk factors for Lyme disease identified with geographic information systems. *Am J Public Health* 85, 944-948.
- Gonzalez C, Wang O, Strutz SE, Gonzalez-Salazar C, Sanchez-Cordero V, Sarkar S, 2010. Climate change and risk of leishmaniasis in North America: predictions from ecological niche models of vector and reservoir species. *PLoS Negl Trop Dis* 4, e585.
- Jensen JR, 2005. Introductory digital image processing: a remote sensing perspective. New York: John Wiley and Sons, 526 p.
- Jerrett M, Gale S, Kontgis C, 2010. Spatial modeling in environmental and public health research. *Int J Environ Res Public Health* 7, 1302-1329.
- Kafadar K, 1994. Choosing among two-dimensional smoothers in practice. *Comput Stat Data Anal* 18, 419-439.

- Kitron U, Kazmierczak JJ, 1997. Spatial analysis of the distribution of Lyme disease in Wisconsin. *Am J Epidemiol* 145, 558-566.
- Kloog I, Haim A, Portnov BA, 2009. Using kernel density function as an urban analysis tool: investigating the association between nightlight exposure and the incidence of breast cancer in Haifa, Israel. *Comput Environ Urban Syst* 33, 55-63.
- Lai PC, Wong CM, Hedley AJ, Lo SV, Leung PY, Kong J, Leung GM, 2004. Understanding the spatial clustering of severe acute respiratory syndrome (SARS) in Hong Kong. *Environ Health Perspect* 112, 1550-1556.
- Landis JR, Koch GG, 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159-174.
- MacEachren AM, Brewer CA, Pickle LW, 1998. Visualizing geo-referenced data: representing reliability of health statistics. *Environ Plan A* 30, 1547-1562.
- Martinez A, Salinas A, Martinez F, Cantu A, Miller DK, 1999. Serosurvey for selected disease agents in white-tailed deer from Mexico. *J Wildl Dis* 35, 799-803.
- National Elevation Dataset, 2010. The national map seamless server. Available at: <http://seamless.usgs.gov> (accessed on February 2010).
- National Land Cover Dataset, 2010. Multi-resolution land characteristics consortium. Available at: www.mrlc.gov (accessed on January 2010).
- Pawar S, Koo MS, Kelley C, Ahmed MF, Chaudhuri S, Sarkar S, 2007. Conservation assessment and prioritization of areas in northeast India: priorities for amphibians and reptiles. *Biol Conserv* 136, 346-361.
- Peterson AT, 2008. Biogeography of diseases: a framework for analysis. *Naturwissenschaften* 95, 483-491.
- Phillips SJ, Anderson RP, Schapire RE, 2006. Maximum entropy modeling of species geographic distributions. *Ecol Model* 190, 231-259.
- Phillips SJ, Dudík M, 2008. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography* 31, 161-175.
- PRISM Climate Group, 2010. Precipitation 2004-2008. Available at: www.prism.oregonstate.edu/index.phtml (accessed on April 2010).
- Rushton G, Peleg I, Banerjee A, Smith G, West M, 2004. Analyzing geographic patterns of disease incidence: rates of late-stage colorectal cancer in Iowa. *J Med Syst* 28, 223-236.
- Sarkar S, Strutz SE, Frank DM, Rivaldi CL, Sissel B, Sanchez-Cordero V, 2010. Chagas disease risk in Texas. *PLoS Negl Trop Dis* 4, e836.
- Serra P, Pons X, Saurí D, 2003. Post-classification change detection with data from different sensors: some accuracy considerations. *Int J Remote Sens* 24, 3311-3340.
- Tiwari C, 2011. Web-based disease mapping and analysis program. Available at: www.webdmap.com (accessed on May 2010).
- Tiwari C, Rushton G, 2005. Using spatially adaptive filters to map late stage colorectal cancer incidence in Iowa. In: *Developments in spatial data handling*. Fischer PF (ed). Berlin: Springer, 665-676 pp.
- U.S. Department of Agriculture, 2010. Soil Survey Geographic (SSURGO) Database. Available at: <http://soils.usda.gov/survey/geography/ssurgo> (accessed on May 2010).
- Williamson PC, Billingsley PM, Teltow GJ, Seals JP, Turnbough MA, Atkinson SF, 2010. *Borrelia, Ehrlichia* and *Rickettsia* spp. in ticks removed from persons, Texas, USA. *Emerg Infect Dis* 16, 441-446.
- WorldClim.org, 2010. Global Climate Data. Available at: www.worldclim.org (accessed on February 2010).
- Yang DH, Goerge R, Mullner R, 2006. Comparing GIS-based methods of measuring spatial accessibility to health services. *J Med Syst* 30, 23-32.